Tentamen Informatietheorie en Codes (kans A)

Opgave 1. *(7 punten)*
Gegeven is de kansverdeling $P = (\frac{8}{23}, \frac{6}{23}, \frac{4}{23}, \frac{2}{23}, \frac{2}{23}, \frac{1}{23})$.

(i) Bepaal de lengtes $l_i = \lceil \log \frac{1}{p_i} \rceil$ van de Shannon-codering voor P.

(ii) Construeer een Huffman-codering voor P, bepaal de gemiddelde codelengte van deze codering en vergelijk deze met de entropie van P.

(iii) Bepaal een kansverdeling P' voor de codewoorden uit de Huffman-codering uit deel (ii), zodat de gemiddelde codelengte bij deze kansverdeling gelijk is aan de entropie $H(P')$.

(iv) Voor de gegeven kansverdeling P geldt dat voor iedere kans p_i de lengte $l_H(p_i)$ in de Huffman-codering kleiner of gelijk is aan de lengte l_i in de Shannon-codering.

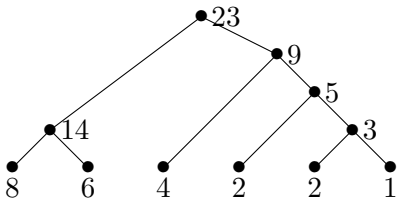
Laat door een tegenvoorbeeld zien, dat dit in het algemeen niet waar hoeft te zijn, d.w.z. vind een kansverdeling $P'' = (p_1, \ldots, p_n)$ zodat in de Huffman-codering van P'' voor een lengte $l_H(p_i)$ geldt dat $l_H(p_i) > \lceil \log \frac{1}{p_i} \rceil$.

Hint: Een kansverdeling op vier waarden is voldoende.

Oplossing:

(i) De lengte voor de Shannon-codering van p_i is l_i als l_i het kleinste getal is zodat $p_i \geq 2^{-l_i}$, of te wel zodat $2^{l_i} \geq \frac{1}{p_i}$. Hiermee vindt men de lengtes $L = (2, 2, 3, 4, 4, 5)$.

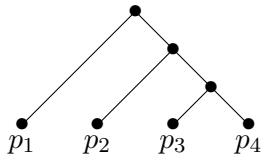
(ii) De boom van de Huffman-codering ziet er als volgt uit:



De lengtes zijn dus $L_H = (2, 2, 2, 3, 4, 4)$ en de gemiddelde codelengte is $\frac{1}{23}(8 \cdot 2 + 6 \cdot 2 + 4 \cdot 2 + 2 \cdot 3 + 2 \cdot 4 + 1 \cdot 4) = \frac{54}{23} \approx 2.348$, terwijl de entropie $H(P) \approx 2.284$ is.

(iii) De gemiddelde codelengte wordt gelijk aan de entropie als alle kansen van de vorm $p_i = 2^{-l_i}$ zijn, dan is $l_i = \log \frac{1}{p_i}$. Met de gevonden lengtes in deel (ii) geeft dit de kansverdeling $P' = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{16})$, hiervoor zijn gemiddelde codelengte en entropie beide $H(P') = 3 \cdot \frac{1}{4} \cdot 2 + \frac{1}{8} \cdot 3 + 2 \cdot \frac{1}{16} \cdot 4 = \frac{3}{2} + \frac{3}{8} + \frac{1}{2} = 2.375$.

(iv) Het idee is dat de Huffman-codering voor $p_1 \geq p_2 \geq p_3 \geq p_4$ een boom van de vorm



oplevert, waar de codering van p_3 lengte 3 heeft, terwijl $p_3 > \frac{1}{4}$. Hiervoor moet alleen $p_3 + p_4 < p_1$ gelden en dit is mogelijk, als p_4 klein is. Een voorbeeld is $p_3 = 0.28$, $p_4 = 0.02$, dan kan $p_2 = 0.30 = p_3 + p_4$ en $p_1 = 0.40$ gekozen worden.

Opgave 2. *(6 punten)*
Laten X, Y en Z stochasten zijn met waarden in $\mathbb{F}_2$, waarbij

$$P(X = 0) = p,\ P(X = 1) = 1 - p = p', \qquad P(Y = 0) = q,\ P(Y = 1) = 1 - q = q'.$$

De stochast Z is de som van X en Y, dus $Z = X + Y$.
Je kunt hierbij X beschouwen als input van een kanaal, Y als ruis en Z als de output van het kanaal.

(i) Bepaal voor het speciale geval $q = \frac{1}{2}$ de kansverdeling $P(Z)$ en de informatie $I(X \mid Z)$ die Z over X onthult.

Licht toe, waarom je het resultaat ook intuïtief had kunnen verwachten.

(ii) Bepaal $P(Z)$ en $I(X \mid Z)$ voor algemene q.

(iii) Stel nu $p = \frac{1}{2}$, d.w.z. de inputs zijn uniform verdeeld. Bepaal ook in dit geval de informatie $I(X \mid Z)$ die Z over X onthult.

Bereken de waarde van deze informatie voor $q = 0.9$.

Oplossing:

(i) $P(Z = 0) = pq + p'q'$ en $P(Z = 1) = pq' + p'q$.

In het speciale geval $q = \frac{1}{2}$ is $P(Z = 0) = P(Z = 1) = \frac{1}{2}$. Hieruit volgt $H(Z) = 1$. Verder is $H(Z \mid X = 0) = H(Z \mid X = 1) = H(\frac{1}{2}, \frac{1}{2}) = 1$ en dus ook $H(Z \mid X) = p \cdot H(Z \mid X = 0) + p' \cdot H(Z \mid X = 1) = 1$ en dus $I(Z \mid X) = I(X \mid Z) = 0$.

Dit zou je ook zo verwachten, want de uniforme verdeling van de ruis maakt alle informatie teniet. Dat Z uniform verdeeld is, is geen toereikend argument, want dat is ook bij deel (iii) het geval.

(ii) De verdeling $P(Z)$ is al in deel (i) bepaald. Hiermee is $H(Z) = H(pq + p'q', pq' + p'q)$ en $H(Z \mid X) = p \cdot H(q, q') + p' \cdot H(q', q) = H(q, q')$ en dus

$$I(X \mid Z) = H(Z) - H(Z \mid X) = H(pq + p'q', pq' + p'q) - H(q, q').$$

(iii) Uit deel (ii) volgt rechtstreeks dat Z uniform verdeeld is en dat dus $H(Z) = 1$, hieruit volgt $I(X \mid Z) = 1 - H(q, q')$.

De waarde voor $q = 0.9$ is $I(X \mid Z) \approx 0.531$ (bits).

Opgave 3. *(10 punten)*

Gegeven is de generator matrix $G_0 = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & -1 \end{pmatrix}$ voor de ternaire $[4,2]$-Hamming code $\mathcal{H}_4$ over $\mathbb{F}_3$. We weten dat $\mathcal{H}_4$ een zelfduale code is met minimum afstand 3.

(i) Je ontvangt de woorden $y_1 = (1,-1,0,0)$, $y_2 = (1,-1,0,-1)$ en $y_3 = (-1,1,-1,1)$.

Decodeer deze woorden middels minimum afstand decodering (dat overeenkomt met maximum likelihood decodering).

(ii) Je verstuurt de codewoorden van de Hamming code $\mathcal{H}_4$ via een *ternair symmetrisch kanaal* met overgangsmatrix $\begin{pmatrix} 1-p & \frac{p}{2} & \frac{p}{2} \\ \frac{p}{2} & 1-p & \frac{p}{2} \\ \frac{p}{2} & \frac{p}{2} & 1-p \end{pmatrix}$.

Via dit kanaal blijft een symbool met kans $1-p$ correct en wordt met kans p veranderd (telkens met kans $\frac{p}{2}$ naar ieder van de twee andere symbolen).

Wat is de kans p_E op een decoderingsfout als de codewoorden van de Hamming code $\mathcal{H}_4$ via dit kanaal verstuurd worden? Bepaal p_E concreet voor $p = 10\%$ en $p = 1\%$.

(iii) De bitrate R van de Hamming code $\mathcal{H}_4$ is natuurlijk $R = \frac{2}{4} = \frac{1}{2}$.

Bepaal de capaciteit van het ternaire symmetrische kanaal voor $p = 10\%$ en laat zien dat deze boven de bitrate van de Hamming code $\mathcal{H}_4$ ligt.

Let op: Omdat we over $\mathbb{F}_3$ werken, moet je in de entropie de logaritme met grondtal 3 gebruiken.

Zij verder I de 4×4-eenheidsmatrix en J de 4×4-matrix met alle elementen 1. Definieer dan de generator matrix G van een lineaire $[12,6]$-code $\mathcal{C}$ over $\mathbb{F}_3$ door

$$G = \begin{pmatrix} J+I & I & I \\ 0 & G_0 & -G_0 \end{pmatrix}$$

(met G_0 de generator matrix van $\mathcal{H}_4$). De matrix G heeft rang 6, omdat $I+J$ rang 4 heeft (want $\det(I+J) = -1$ over $\mathbb{F}_3$).

(iv) Laat zien dat $\mathcal{C}$ een zelfduale code is.

(v) Bewijs dat voor iedere lineaire zelfduale code over $\mathbb{F}_3$ geldt, dat de minimum afstand d een veelvoud is van 3.

(vi) Ga na dat $\mathcal{C}$ minimum afstand 6 heeft.

Oplossing:

(i) Omdat $\mathcal{H}_4$ zelfduaal is, is $H = G_0$ een parity-check matrix voor $\mathcal{H}_4$. Verder weten we dat $\mathcal{H}_4$ een 1-foutverbeterende perfecte code is, dus is een ontvangen woord of een codewoord of heeft afstand 1 van een codewoord. De locatie van de fout kunnen we middels syndroom decodering achterhalen. Er geldt $y_1 H^{tr} = (1,-1)$ en dit is de vierde kolom van H, dus is $y_1 = c + e_4$ voor een codewoord c, dus decoderen we y_1 als $c = y_1 - e_4 = (1,-1,0,-1) \in \mathcal{H}_4$. Verder is $y_2 H^{tr} = (0,0)$, hier is dus geen fout opgetreden en we decoderen y_2 als y_2. Ten slotte is $y_3 H^{tr} = (-1,-1)$, en dit is het negatieve van de derde kolom van H, dus is $y_3 = c - e_3$ en we decoderen y_3 als $c = y_3 + e_3 = (-1,1,0,1)$.

(ii) We decoderen correct als geen of één fout optreden, de kans hierop is $(1-p)^4+4p(1-p)^3$, dus is $p_E = 1-(1-p)^4-4p(1-p)^3$. Voor $p = 10\%$ geeft dit $p_E = 5.23\%$ en voor $p = 1\%$ vinden we $p_E \approx 0.06\%$.

(iii) Omdat dit een symmetrisch kanaal is (iedere rij van de overgangsmatrix is een permutatie van de eerste rij, iedere kolom een permutatie van de eerste kolom), is de capaciteit gegeven door $C = {}^3\log 3 - H_3(1-p, \frac{p}{2}, \frac{p}{2}) = 1 + (1-p)\,{}^3\log(1-p) + 2\cdot\frac{p}{2}\,{}^3\log\frac{p}{2}$. Voor $p = 0.1 = 10\%$ geeft dit $C \approx 0.641$ en dit ligt boven de bitrate $R = \frac{1}{2}$.

(iv) Merk op dat $J^2 = J$, hieruit volgt $(J+I)(J+I)^{tr} = J^2 + 2J + I = J + 2J + I = I$. We zien dus dat

$$G\,G^{tr} = \begin{pmatrix} J+I & I & I \\ 0 & G_0 & -G_0 \end{pmatrix}\begin{pmatrix} J+I & 0 \\ I & G_0^{tr} \\ I & -G_0^{tr} \end{pmatrix} = \begin{pmatrix} I+I+I & G_0^{tr}-G_0^{tr} \\ G_0-G_0 & G_0G_0^{tr}+G_0G_0^{tr} \end{pmatrix}$$
$$= \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix},$$

waarbij $G_0G_0^{tr} = 0$ geldt omdat $\mathcal{H}_4$ zelfduaal is.

(v) Voor een codewoord $c = (c_1, \ldots, c_n)$ over $\mathbb{F}_3$ is $\langle c, c\rangle = \sum_{i=1}^n c_i^2 = \sum_{c_i\neq 0} 1$, want $1\cdot 1 = (-1)\cdot(-1) = 1$ in $\mathbb{F}_3$. Dus is $\langle c, c\rangle = w(c) \bmod 3$ en voor een zelfduale code is dan $w(c) \equiv 0 \bmod 3$, omdat in dit geval $\langle c, c\rangle = 0$ voor alle $c \in \mathcal{C}$. Omdat de minimum afstand gelijk is aan het minimale gewicht van een niet-nul vector, is de minimum afstand een veelvoud van 3.

(vi) Iedere van de rijen van G heeft gewicht 6, dus is de minimum afstand hoogstens 6. Volgens deel (v) zijn we klaar als we kunnen aantonen dat iedere lineaire combinatie van de rijen van G minstens gewicht 4 heeft. Alternatief zouden we ook kunnen aantonen dat iedere combinatie van drie kolommen van G lineair onafhankelijk is, want dan is de minimum afstand groter dan 3.

Het is duidelijk dat iedere lineaire combinatie van de laatste twee rijen van G minstens gewicht 6 heeft, en iedere lineaire combinatie van twee of meer van de eerste vier rijen heeft al in de laatste acht coordinaten minstens gewicht 4. Als we alleen naar de laatste acht coordinaten kijken (de laatste twee blokken), zien we dat optellen van een vector van de vorm (v, v) (dus een lineaire combinatie van rijen uit (I, I)) bij een vector van de vorm $(w, -w)$ (een lineaire combinatie van de laatste twee rijen) alleen in w of in $-w$ een coordinaat tot nul kan maken, maar niet in beide. Hierdoor is het gewicht van een lineaire combinatie in de laatste acht coordinaten minstens 3. Maar in de eerste vier coordinaten komt dan nog minstens gewicht 1 erbij, want $J+I$ is inverteerbaar en iedere niet-triviale lineaire combinatie is niet-nul.